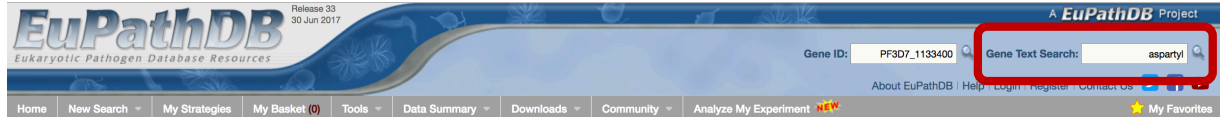


Finding Genes, Building Search Strategies and Visiting a Gene Page

1. Finding genes that encode aspartyl proteases using text search.
For this exercise navigate to <http://eupathdb.org>

- a. Find all genes that contain the word aspartyl in the gene record.

Hint: use the keyword aspartyl “Gene Text Search” box.



- How many genes did you get?
- Look closely at the different sections of the result page. How many of those are in *Toxoplasma gondii* strain GT1? How did you find this out?
(Hint – the filter table is located between the strategy panel and the result table and shows the distribution of results across the organisms that you searched. Click on a number to only display results from a specific species or strain – you will have to scroll to the right of the page since the table is wide).

[illegible]

- b. Filter your results to only look at the results from *T. gondii* GT1.

To filter your results, click on the number in the box under GT1

										Chromerida										Diplomonadida		
Sarcocystis					Theileria					Toxoplasma										Giardia		
Pyoelii (nr Genes: 38)					S.neurona (nr Genes: 15)					T.gondii (nr Genes: 119)										C.vella		
T.annulata					T.equ					T.gondii										G.Assemblage A2		
strain Ankara					strain WA					strain Shintoku										isolate WB		
strain Muguga					strain Muguga					strain Muguga										isolate GS_B		
ARI					FOU					GAB2-2007-GAL-DOM2										CCMP2878		
GT1					MAS					ME49										CCMP3155		
RH					RUB					TgCatPRC2										isolate DH		
VAND					VEG					p89										isolate WB		
15					11					10										3		
12					10					12										3		
10					7					7										3		

Gene Results

Genome View

Analyze Results

Advanced Paging

Download

Add to Basket

Add Columns

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	Found in	Score
TGGT1_242720	TGGT1_242720-126_1	T. gondii GT1	TGGT1_chrVII:2,116,490..2,124,101(-)	aspartyl protease ASP5	User Comments, InterPro, Product	25
TGGT1_209620	TGGT1_209620-126_1	T. gondii GT1	TGGT1_chrib:1,378,624..1,380,201(-)	eukaryotic aspartyl protease superfamily protein	InterPro, Product	2
TGGT1_201840	TGGT1_201840-126_1	T. gondii GT1	TGGT1_chrVIIa:3,648,103..3,654,235(+)	aspartyl protease ASP1	InterPro, Product	2
TGGT1_262940	TGGT1_262940-126_1	T. gondii GT1	TGGT1_chrVIIb:881,585..883,991(+)	putative aspartyl proteinase (eimepsin)	InterPro, Product	2
TGGT1_272510	TGGT1_272510-126_1	T. gondii GT1	TGGT1_chrVIII:3,925,164..3,927,417(-)	aspartyl protease	InterPro, Product	2
TGGT1_246550	TGGT1_246550-126_1	T. gondii GT1	TGGT1_chrXII:2,777,006..2,782,288(+)	aspartyl protease ASP3	InterPro, Product	2
TGGT1_297970	TGGT1_297970-126_1	T. gondii GT1	TGGT1_chriII:2,122,080..2,130,718(+)	aspartyl aminopeptidase	Product	1
TGGT1_304680	TGGT1_304680-126_1	T. gondii GT1	TGGT1_chrVIIa:595,893..600,165(-)	ubiquitin family protein	InterPro	1
TGGT1_202530	TGGT1_202530-126_1	T. gondii GT1	TGGT1_chrVIIa:3,153,086..3,158,845(-)	aspartate-tRNA ligase	GOTerms	1
TGGT1_261530	TGGT1_261530-126_1	T. gondii GT1	TGGT1_chrVIIb:1,638,349..1,643,797(-)	eukaryotic aspartyl protease superfamily protein	Product	1

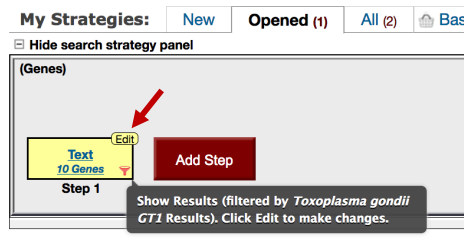
Notice how the number in the box become highlighted in yellow.

c. Examine your results. Specifically look at the Product Description column.

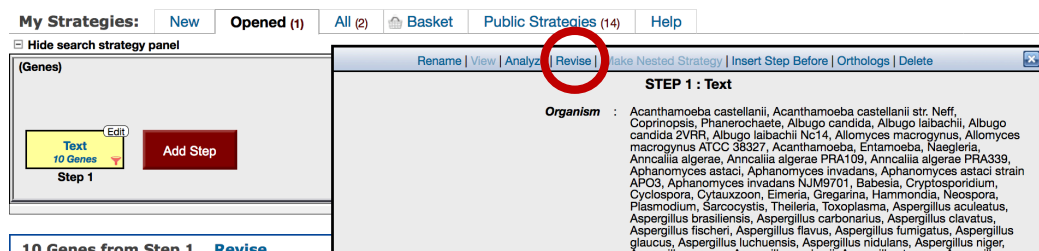
How many genes have the word aspartyl in their description?

d. Revising a search.

Revising a search allows you to change the parameters of a search. To revise a search move your cursor over the box in your search strategy until you see the edit link.



Click on the edit link then click on the “Revise” link



Once you click on revise, on the next popup page, scroll to the “Fields” section and only choose the field called “Gene Product”

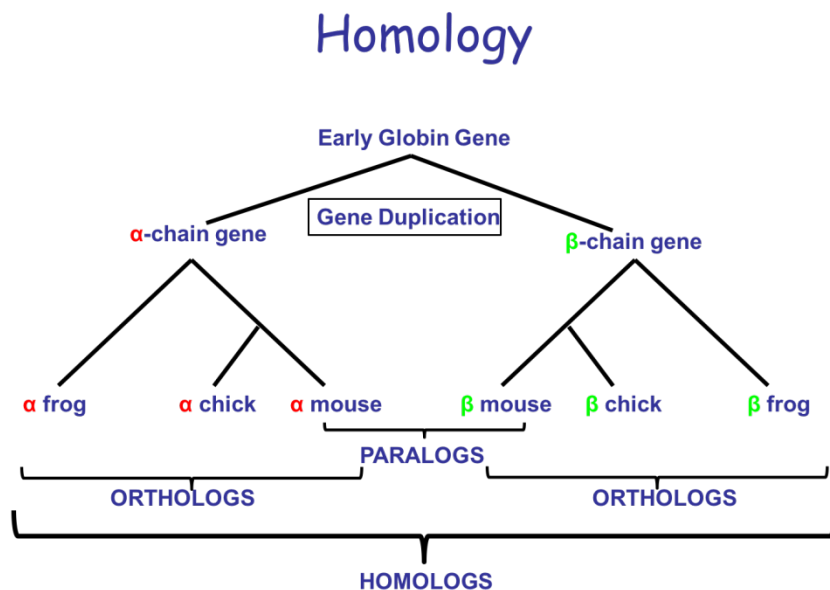
Fields ?

- ☐ Alias
- ☐ EC descriptions
- ☐ Gene ID
- ☐ Gene notes
- ☒ Gene product
- ☐ Gene name
- ☐ GO terms and definitions
- ☐ Metabolic pathway names and descriptions
- ☐ Phenotype
- ☐ Protein domain names and descriptions
- ☐ PubMed
- ☐ Rodent Malaria Phenotype
- ☐ Similar proteins (BLAST hits v. NRDB/PDB)
- ☐ User comments

[select all](#) | [clear all](#)

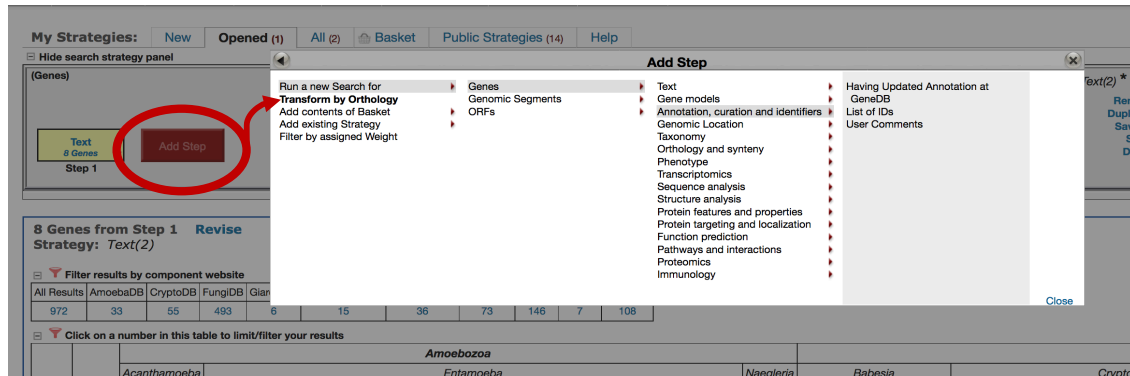
Next click on the “Run Search” button. How did your results change? Look at the product description column – do all genes have the word aspartyl in their descriptions?

2. Leveraging orthology to identify more aspartyl proteases (we will have a discussion on orthology)

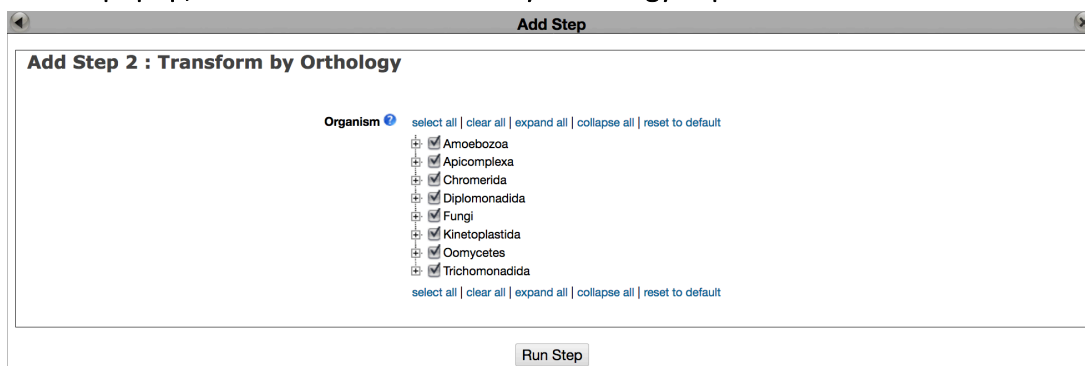


EuPathDB resources have a tool that allows you to transform a list of genes to their orthologs and paralogs across organisms in the database.

- Starting with your results in part 1, click on the add step button (red button in the strategy panel).



- In the popup, click on the “Transform by Orthology” option.



Keep the default parameters selected and click on the Run Step button. Think about what are you asking the database to return.

- What do your results look like? How many genes in total did you get? What is the distribution of orthologs across apicomplexan (hint: explore the filter table)?
- Filter your results to view genes from *Plasmodium falciparum* 3D7. What are the genes called?

											Plasmodium										
Gregarina			Hammondia		Neospora																
Is	E.necatrix	E.praecox	E.tenella	G.niphandroses	H.hammondi	N.caninum	P.berghiei	P.chabaudi	P.coatneyi	P.cynomolgi	P.falciparum (n: 20)	P.fragile	P.gaboni	P.gallinaceum	P.indu	P.knowlesi	P.malariae	P.ovale curtsii	P.reichenow	P.prellii	
iton	Houghton	Houghton	strain Houghton	Unknown strain	strain H.H.34	Liverpool	ANKA	chabaudi	Hackeri	strain B	3D7	IT	strain nigrit	strain SY75	8A	San Antonio 1	strain H	UG01	GH01	CDC	SGS
<div><div>6</div><div>2</div><div>6</div><div>6</div><div>8</div><div>8</div><div>8</div><div>8</div><div>7</div><div>8</div><div>10</div><div>7</div><div>10</div><div>8</div><div>7</div><div>8</div><div>9</div><div>8</div><div>10</div><div>8</div></div>																					

Gene Results

Genome View

Analyze Results

Advanced Paging

Gene ID

Transcript ID

Organism

Genomic Location (Gene)

Product Description

Input Ortholog(s)

Ortholog Group

Paralog count

Ortholog count

PF3D7_0311700

PF3D7_0311700.1

P. falciparum 3D7

Pf3D7_03_v3:502,698..505,848(+)

plasmeprin VI

TGGT1_201840,
TGGT1_262940

OG5_126885

3

46

PF3D7_1407800

PF3D7_1407800.1

P. falciparum 3D7

Pf3D7_14_v3:283,086..284,435(+)

plasmeprin IV

TGGT1_201840,
TGGT1_262940

OG5_126885

3

46

PF3D7_1407900

PF3D7_1407900.1

P. falciparum 3D7

Pf3D7_14_v3:288,297..289,655(+)

plasmeprin I

TGGT1_201840,
TGGT1_262940

OG5_126885

3

46

PF3D7_1408000

PF3D7_1408000.1

P. falciparum 3D7

Pf3D7_14_v3:293,471..294,832(+)

plasmeprin II

TGGT1_201840,
TGGT1_262940

OG5_126885

3

46

PF3D7_0932300

PF3D7_0932300.1

P. falciparum 3D7

Pf3D7_09_v3:1,289,315..1,291,027(-)

M18 aspartyl
aminopeptidase

TGGT1_297970

OG5_128020

0

21

PF3D7_1323500

PF3D7_1323500.1

P. falciparum 3D7

Pf3D7_13_v3:975,403..977,175(+)

plasmeprin V

TGGT1_242720

OG5_132039

0

21

PF3D7_0808200

PF3D7_0808200.1

P. falciparum 3D7

Pf3D7_08_v3:416,344..418,065(-)

plasmeprin X

TGGT1_246550

OG5_133546

1

43

PF3D7_1430200

PF3D7_1430200.1

P. falciparum 3D7

Pf3D7_14_v3:1,188,349..1,191,466(+)

plasmeprin IX

TGGT1_246550

OG5_133546

1

43

PF3D7_1033800

PF3D7_1033800.1

P. falciparum 3D7

Pf3D7_10_v3:1,351,197..1,353,284(+)

plasmeprin VII

TGGT1_272510

OG5_145118

0

21

PF3D7_1465700

PF3D7_1465700.1

P. falciparum 3D7

Pf3D7_14_v3:2,658,255..2,660,911(-)

plasmeprin VIII, putative

TGGT1_209620

OG5_150586

0

21

3. Exploring specific aspartyl protease gene.

The column called “Gene ID” contains unique gene identifiers for each of the genes in your list. These IDs are also linked to the gene page.

- Click on the gene ID for plasmepsin IV (ID: PF3D7_0311700).
- Notice you are now on the gene page for plasmepsin IV in PlasmoDB
- Try and determine at which stage of the plasmodium life cycle is this gene expressed. To do this go to the transcriptomics section of the gene page (hint: you can click on the transcriptomics short cut at the top of the page or you can scroll down the page until you find this section).

[Add to basket](#) [Add to favorites](#) [Download Gene](#)

PF3D7_0311700 plasmepsin VI

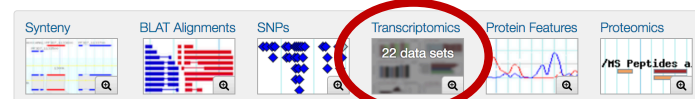
Type: protein coding
Chromosome: 03
Location: PF3D7_03_v3:502,698..505,848(+)

Species: *Plasmodium falciparum*
Strain: 3D7
Status: **Curated** Reference Strain

[View updated annotation at GeneDB](#)
[View 1 user comment, or add a comment](#)

GeneDB curates, researches and improves this genome, and will incorporate appropriate User Comments into the official annotation. If you wish to publish whole genome or large-scale analyses, please contact the primary investigator or use the published version in the PlasmoDB version 5.3 download folder.

Shortcuts



Also see PF3D7_0311700 in the [Genome Browser](#) or [Protein Browser](#)

[Contents](#) << 1 **Gene models** §

[Expand All](#) | [Collapse All](#)

- For example look at the first experiment in the list “Erythrocytic expression time series (3D7, DD2, HB3)”. Is this gene expressed in this time series?
- Now, examine the experiment called “Strand specific transcriptomes of 4 life cycle stages”. Is this gene expressed at the stages represented in this experiment? Which stages?
- Repeat this for each of the aspartyl proteases in *Plasmodium*. Are these proteases expressed at different stages?